JSC-08450

NASA CR-
140221

EARTH RESOURCES PROGRAM

DEVELOPMENT OF A TWO-CHANNEL LINEAR DISCRIMINANT FUNCTION

FOR DETECTING AND IDENTIFYING SURFACE WATER

USING ERTS-1 DATA

SCIENCE AND APPLICATIONS DIRECTORATE

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

AUGUST 1974
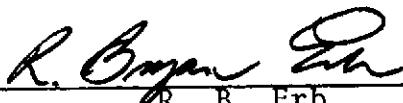
DEVELOPMENT OF A TWO-CHANNEL LINEAR DISCRIMINANT FUNCTION FOR DETECTING AND IDENTIFYING SURFACE WATER USING ERTS-1 DATA

Approved By:

R. B. Erb
Manager, Applications Office

R. B. MacDonald
Chief, Earth Observations Division

EARTH OBSERVATIONS DIVISION
SCIENCE AND APPLICATIONS DIRECTORATE
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
Houston, Texas

August 1973

PREFACE

The purpose of this document is to describe the development of a two-channel linear discriminant function for detecting and identifying surface water using ERTS-1 data. The document presents the background information which guided the development; the characteristics of the input data; and the pertinent characteristics of the computer software. This document includes a discussion of the spectral characteristics of water, and how these spectral properties relate to the two-channel discriminant function for detecting and locating surface water using the ERTS-1 data. The procedures utilized in developing the function are discussed in detail and the constraints associated with the operational use of the discriminant function are presented.

This document was prepared pursuant to requirements identified within the Applications Office of the Earth Observations Division. It is the result of a joint effort by personnel within the Earth Observations Division and personnel within the Earth Resources Applications Department, Lockheed Electronics Company, Inc. Prime technical contributions to the effort described herein were made by A. C. Anderson, Principal Engineer, LEC. Activities by the contractor were authorized under Contract NAS 9-12200.

CONTENTS

LIST OF FIGURES

ACRONYMS

DISPLAY   A LARSAA processor which allows a user to threshold pixels (i.e. assign a pixel to the unclassified category if its likelihood does not exceed a user-specified confidence level) and display classification results in the form of a line printer character map.

EOD   Earth Observations Division

ERTS-1   Earth Resources Technology Satellite (launched July 1972).

IR   Infrared

ISOCLS   An Iterative Self-Organizing Clustering Program which groups spectrally "similar" pixels into sets called clusters.

JSC   Johnson Space Center

LARSAA   A computer software system used to classify and display multispectral scanner data on the basis of spectral information.

MSS   The multispectral scanner on board the Earth Resources Technology Satellite.

PICMON   A processor in LARSAA which generates a grey scale map on a line printer from a single channel of multispectral scanner data.

NDPF   NASA Data Processing Facility

NASA   National Aeronautics and Space Administration

TWRC   Texas Water Rights Commission

CCT   Computer Compatible Tape

ix

1.0 SUMMARY

This document describes the development of a method for detecting surface water using ERTS-1 multispectral scanner digital data. The several computer programs for processing multispectral data which were available are described, along with the manner in which these programs were used in this effort. The use of the LARSAA program to test and demonstrate the discriminant function is explained in detail. The discriminant function itself was later implemented in an operational program which performs only those specific operations necessary to test each data point for an indication of surface water in the scene. Several possible sources of difficulty in the use of the function are identified and suggestions are made for avoiding or correcting those difficulties.

Initial work in the EOD ERTS-1 Project[1] revealed that the range of data values for water in east Texas were lower than for other sur- face features of the scene in all four ERTS-1 MSS spectral channels. Work with several east Texas ERTS-1 scenes recorded on different dates confirmed this fact and indicated that the spectral values for water are fairly consistent. That is, they do not change appreciably from season to season for a given body of water. However, they do change as a function of water turbidity. In addition, pixels[2] related to

---

[1]An EOD effort to analyze ERTS-1 data to assess its ability for detecting and identifying features of applications interest (Reference A Compendium of ERTS-1 Investigations - JSC-08455).

[2]A pixel (picture element) is the basic unit in image reconstruct- ion from the digital tape, using electronic display devices. The intensity value assigned to the element corresponds to the binary integer recorded on digital tape that represents a time sample of the analog scan line trace, the value of which is proportional to the energy sensed in each ERTS-1 MSS channel.

geographical areas along the edges of lakes have energy responses
different from those in the middle of lakes due to inclusion of other
elements of the scene in the instantaneous field of view of the scanner.
The turbidity and the edge pixel discrimination problems were addressed
simultaneously by placing a dividing line in the two channel spectral
space defined using ERTS-1 channels 1 and 4. This line separates water
from non-water and includes in the water class the deep clear water,
the turbid water and the edge pixels which contain a high percentage
of water. The pixels which contain lower percentages of water are
included in the non-water class because the high proportion of back-
ground (non-water area) increases the spectral response levels in
relation to water classes.

Constraints related to low sun elevation angles, cloud shadows
and terrain shadows are discussed briefly in this document and possible
ways of addressing these constraints or correcting the discriminant
function to account for these parameters are outlined.

2.0 INTRODUCTION

In August 1972, the President signed into law Public Law 92-367 authorizing the Secretary of the Army to undertake a national program of inspection of dams. The need for dam safety was brought to national attention when water impoundments in West Virginia and South Dakota gave way, resulting in significant loss of life and property.

In brief, the law directs the Secretary of the Army, acting through the Chief of Engineers, to carry out a national program for the inspection of dams. To determine whether a dam (including the waters impounded by such dam) constitutes a danger to human life or property, the Secretary shall take into consideration the possibility that the dam might be endangered by overtopping, seepage, settlement, erosion, sediment, cracking, earth movement, earthquakes, failure of bulkheads, flashboard, gates on conduits or other conditions which exist or which might occur in any area in the vicinity of the dam.

The report by the Secretary of the Army to Congress is due on or before July 1, 1974. The report shall include (1) an inventory of all dams located in the United States, (2) a review of each inspection made, the recommendations furnished to the governor of the state in which such dam is located and information as to the implementation of such recommendations, and (3) recommendations for a comprehensive national program for the inspection and regulation for safety purposes of dams of the nation, and the respective responsibilities which should be assumed by the federal, state, and local governments and by public and private interest.

In December 1972, the Texas Water Rights Commission
(TWRC) submitted through the Office of the Govenor of Texas,
a request for assistance by NASA/JCS/EOD in the development
of a procedure or procedures for utilizing data acquired by
the Earth Resources Technology Satellite (ERTS-1) in detecting
and locating water impoundments. In response to this request,
a procedure was developed for detecting and locating surface
water using ERTS-1 MSS data. This report traces the develop-
ment of a discriminant function, and presents conclusions and
constraints related to incorporation of this function into
the requested procedure.

Performance criteria were jointly established by the
TWRC and EOD to provide a basis for evaluating the procedure.
These criteria required that areas of surface water ten
acres or greater be correctly identified with an accuracy of
90 percent or greater, with a false detection frequency of
10 percent or less.

ERTS-1 data first became available to the Earth Obser-
vations Division in the late summer of 1972. One of the first
characteristics of the data that was noted was the low rela-
tive magnitude between the data values for water in comparison
to other elements in the scene. This characteristic was noted
in all spectral channels. The spectral signatures of various
targets, particularly water, were extracted and their behaviors
as functions of time, location, season, turbidity and atmos-
pheric condition were noted.

It was determined that the spectral signature for water
is relatively stable over time, and that it is reasonably well

separated in spectral space from other materials found in the rest of the scene. A number of techniques were tested and these activities reported in a companion document (Development of a Computer-Aided Procedure for the National Program of Inspection of Dams, JCS-08449). The present work describes an approach based on establishing a water-non-water boundary in spectral space. Experience gained using ERTS-1 data indicated that the use of a simplified algorithm using data from only two of the ERTS-1 channels had great potential as a candidate for use in an operational program. These considerations ultimately led to the use of a linear discriminant function in two dimensional spectral space (ERTS-1 channels 1 and 4) to discriminate water from all other features contained within the scene.

This discussion begins with some of the earliest assessments of ERTS-1 multispectral scanner (MSS) data and reports the development of a water discriminant function which has the potential for incorporation into an operational data processing software package.

3.0 DATA AND COMPUTER SOFTWARE CHARACTERISTICS

It is important to have an understanding of ERTS-1 data and the computer-aided procedures used for analyzing it before attempting to understand how the linear discriminant function was developed. Consequently, this section discusses the ERTS-1 data and the pertinent computer software programs before addressing the spectral signature of water and the development of the linear discriminant. An understanding of the development of the linear discriminant function is aided by common vocabulary and a geometric interpretation of clustering and classifying data in the two dimensional spaces generated by selected pairs of the four channels of ERTS-1 spectral data.

3.1 ERTS-1 Data

The ERTS-1 MSS data are available on computer compatible digital tapes. A single tape contains the data for an area on the ground which is 25 nautical miles (n.mi.) wide and 96.3 n.mi. long. A scene corresponding to the film image consists of four digital tapes giving an area approximately 100 n.mi. by 96.3 n.mi. The data are 8 bit binary integers with 7 bits being used for channels 1, 2, and 3 (data range 0 to 127) and 6 bits being used for channel 4 (data range 0 to 63).

A calibration is applied to the data by the NASA Data Processing Facility to make the data proportional to the measured scene radiance (brightness) in each channel. The scene radiance which produces a full scale reading is different for each channel, but the data from each channel is linearly related to scene radiance in its particular spectral band. The atmosphere scatters energy in the visible region of the spectrum so there is a "path radiance" which is always present regardless of the scene radiance. Only in the infrared channel 4 is the path radiance low enough that in practice a zero reading can be obtained.

3.2   Computer Processing Software

Several computer programs were available for studying various approaches to detection of surface water. The normal statistical considerations upon which these programs are based were not emphasized in this study. Instead, the positional relationships between cluster centers in spectral space and the curves of constant distance and the geometry of the interclass boundary curves were emphasized. For instance, the interclass boundary is normally interpreted as the locus of points where the probabilities of a data sample being in either of two classes are equal. However, it is also the locus of points where the distances (not necessarily Euclidean distances) from the means of the two classes are equal. In addition, it is a curve (or surface) which separates one class from another. It is this third interpretation which was fundamental to the development and testing of the discriminant function described in this document.

The pertinent features of the programs used, along with the geometric interpretation of their operation, are included in this section as background for the discussion to follow. The programs which were used to develop and test the discriminant function were PICMON, ISOCLS, and LARSAA.

3.2.1   PICMON Program

The PICMON processor outputs the ERTS-1 data for a single channel as a line printer map. The program has 16 bins (assignable data range intervals) that it can represent with 16 printer symbols. The bin edges[3] can be determined by the program using the histogramming option,

---

[3]Bin edges were defined as the upper and lower limits of the range of data values which are to be included in the bin.

or they can be specified by the analyst. The printer symbols can be the set automatically assigned by the program or they can be specified by the analyst. The most useful option for investigation the spectral signature of water for this effort consisted of setting the bin edges so that each bin contained only one integer. In this manner, the exact data value could be seen and the location in the line printer map of data values which differed from the average could be readily found. Experience gained during data analysis within the EOD ERTS-1 project indicated that large, clear bodies of water could be located by displaying channel 4 data values from zero to five and representing all higher values as blanks. In this effort the PICMON program was used to print water grey maps by setting the first six bins of channel 4 data as the integers 0 through 5 and by specifying the printer symbol "blank" for all bins above the bottom six. These grey maps were generated to delineate large clear bodies of water.

3.2.2 ISOCLS Program

The ISOCLS processor groups spectrally similar pixels into clusters. It accomplishes this by matching the data levels in all specified channels to the mean vectors in each of an existing set of clusters. Each new pixel considered against all existing clusters and is then assigned to the one to which it is most similar. The similarity is measured by forming the ℓ_1 distance to each cluster in the spectral space made up of the data in the scanner channels and assigning the pixel to the cluster which is closest. The ℓ_1 distance is the sum of the absolute differences between the pixel data values and the cluster mean vector values.

The geometric interpretation of the ℓ_1 distance measure can be pictured in a two dimensional spectral space to obtain a feel for how it behaves. The curve of constant ℓ_1 distance from a point in two dimensional space is a diamond whose edges cut the coordinate axes at

at 45°. The curves which separate clusters are straight lines. They are oriented so that they cut the coordinate axes at either 0°, 45°, or 90°. No other orientations are allowed. An example of intercluster boundaries using the ℓ_1 distance measure is shown in Figure 1.

If the data actually occur in a long, narrow cluster, the cluster can still be contained within the diamond which encloses the mean vectors but, if the diamond is allowed to be too small, the ends of the cluster will be cut off and the pixels will be assigned to another cluster.

Earlier EOD ERTS-1 Project uses of the ISOCLS program in the study of spectral signatures of water met with some difficulties. Water is quite well separated in spectral space from other materials contained within the scene. Large, clear bodies of water are contained in one cluster. Since it was necessary to distinguish between water areas of differing turbidities, the allowable size of the clusters was decreased in an attempt to separate water containing various levels of turbidity. The result was to form a large number of clusters in the regions of spectral space occupied by non-water targets without forming additional clusters in the region of spectral space occupied by water. Since the water had low data values in all four channels and smaller clusters were required to discriminate the water containing several levels of turbidity, a modification was made to the ISOCLS program to permit allowable cluster dimensions in spectral space to be smaller for smaller data values. Several possibilities were considered. The one implemented required that the standard deviation in each channel in each cluster be less than a specified constant times the square root of the data value in that channel. It was determined that 0.8 produced a number of clusters in the region of spectral space occupied by water pixels, which represented water of different turbidities. This modified ISOCLS program was used within this effort.
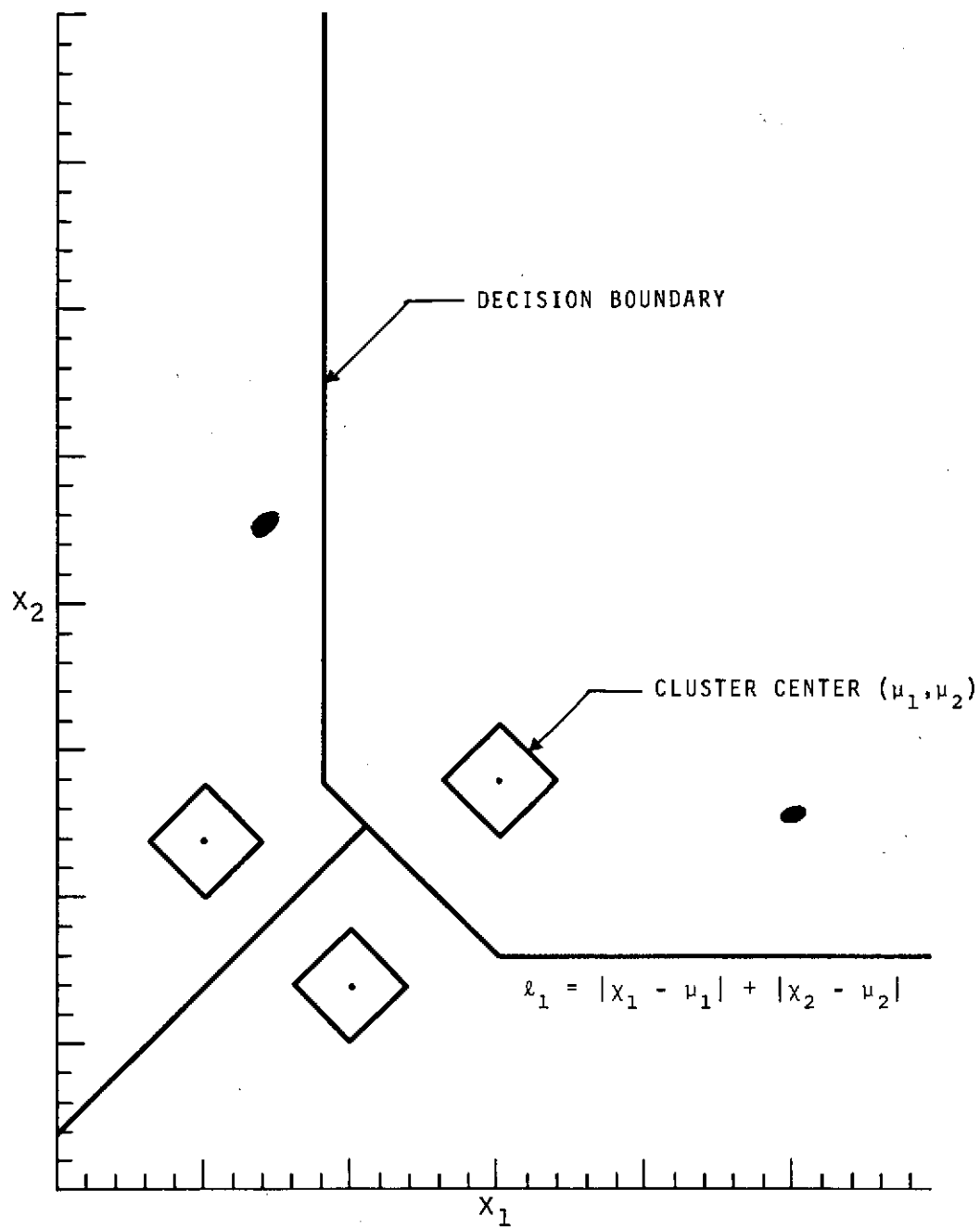
Figure 1. — Illustration of the use of the ℓ_1 distance measure. The diamonds are at a distance of two units from the cluster centers. The intercluster boundaries are equidistant from two cluster centers.

3.2.3 LARSAA Program

The LARSAA program contains three processors used within this effort. The STAT processor takes data for designated training fields and calculates the channel means and covariances for each training field. It provides the statistics for the CLASSIFY processor which assigns each pixel to one of the classes which has been identified to the STAT processor by the input of specific training fields. CLASSIFY uses a quadratic distance measure (d) to determine to which class a pixel is closest. This distance measure is given by the square root of a quadratic form as follows:

$$d^2 = (X - \mu_i)^T K_i^{-1} (X - \mu_i)$$

where

    X is the vector of data values for a pixel

    μ_i is the vector of mean values for class i

    K_i^{-1} is the inverse of the covariance matrix for class i

    d is the CLASSIFY quadratic distance.

The curves of constant distance from a class mean in two dimensional space are ellipses. The ellipses may be oriented in any direction in the spectral space. When the off-diagonal elements of the covariance matrix are zero, the ellipses are oriented with their axes parallel to the coordinate axes. When the off-diagonal elements are zero and the diagonal elements are equal, the curves of constant distance are circular and the distance measure reduces to a scaled Euclidean distance. The boundary between classes is the locus of points which are equidistant from two class means. They are quadratic curves in two dimensional spectral space. The distance measure is different in each class and equating the distances from two classes is not equivalent to equating Euclidean distances. The components of the distance measures are normalized to the square roots of the diagonal

3-6

elements in the covariance matrix. If the normalizing factors are small for one class and large for another, the class with the large normalization factors will occupy the greater part of the spectral space.

The LARSAA software partitions the spectral space into a number of classes, separated by quadratic boundaries. The matrices which generate the normalized distance measures which, in turn, generate the quadratic boundaries, can come from the STAT processor or may be generated independently and input by the analyst.

The third processor in LARSAA is DISPLAY, which produces a line printer map of the classified area using a specified printer symbol for each class. This processor also has the option of printing a blank if the data vector for a pixel is more than a specified normalized distance from its class mean.

4.0 DEVELOPMENT OF A WATER DISCRIMINANT TECHNIQUE

The computer programs discussed in the previous section were used to develop and test a discriminant function for detection of surface water. The manner in which they were used to accumulate the necessary data and to generate the water discriminant function is described in this section.

4.1 Spectral Signature of Water

Experience gained during analysis within the EOD ERTS-1 project indicated that the channel 4 data values for water were generally in the range of zero to 5. PICMON grey maps were generated to determine which additional elements of the scene began to appear as the higher data values in channel 4 were allowed to increase. Using channel 4 data values of 6 to 9, pixels around the perimeter of large areas of surface water began to appear, in addition to a few isolated pixels away from the large areas of surface water. Using data values of 10 to 12 in channel 4, additional perimeter pixels appeared, and some of the isolated pixels became isolated groups. The isolated groups were identified as small areas of surface water when correlated to aircraft photography of the area.

A study was initiated to relate pixels classified as water to areas of surface water of various sizes. Two goals were identified: (1) to determine the size of the smallest area of surface water detectable; and, (2) to determine if there was a relation between the area of the surface water and the number of pixels with low data values in channel 4. The extent of several hundred areas of surface water were measured from aerial photography to determine the probability of detection as a function of area of surface water. It should be emphasized that evaluation of the frequency of false alarms was not attempted in these earlier studies. However, a sufficient number of selected areas of surface

water were identified and correlated with the computer printout to indicate that approximately 50 percent correct detection occurred for areas of surface water of about one acre. Larger surface water areas had higher frequencies of correct detection. This correlation is a function of the finite size of the MSS sensor field-of-view and the response time of the scanner. When an area of surface water is smaller than the sensor instantaneous field-of-view, the scanner sees a mixture of water and some other surface material. Using only channel 4 data with a range of from 0 to 13, the detection probability was essentially 100 percent for areas of surface water larger than two acres.

4.2 Development of a Computer-Aided Procedure

The effort in EOD related to the National Program of Inspection of Dams required further study of the possibility of using ERTS-1 data to detect and locate surface water. The EOD ERTS-1 Project results discussed above were used as a starting point and several procedures were evaluated to determine their usefulness for detecting and locating surface water. These procedures are described in detail in "Development of a Computer-Aided Procedure for the National Program of Inspection of Dams", JSC-08449.

A study area was selected which encompassed portions of Austin, Brazos, Burleson, Colorado, and Washington Counties in Texas (ERTS-1 scene E-1092-16305, 23 October 1972). Ground truth information was extracted from RB57F aerial color infrared photography at a scale of 1:120,000 (Mission 220, flown 8 November 1972). This area includes Lake Somerville and numerous smaller bodies of water of various sizes. A channel 4 grey map of the study area was produced using the PICMON program. The low data values (0 to 9), which experience had shown to be associated with water, were printed to display pixels that represented areas of surface water. Large areas of surface water such as Lake Somerville and medium sized areas of water were easily identified because

the channel 4 data values were below 7. The scene from the 23 October 1972 satellite pass had non-water data values lower than those for earlier scenes, primarily because the sun elevation of 42° was below those for earlier scenes. Preliminary analyses indicate that data values for water are less sensitive than data values for non-water to sun elevations in the range of 30° to 60°. The October 23 scene was acquired when recent rains had caused substantial amounts of standing water in the study area. This condition represented a worst case situation for separating areas of persistent surface water from areas which contained a mixture of small amounts of standing water plus other surface materials.

The PICMON grey map was used to select training fields for "water only" as input to the LARSAA - STAT processor. The bulk of the training data was selected from Lake Somerville. Several hundred pixels in one large block were picked from the center of the lake. They outweighed the small pond data in the generation of the class mean vector in LARSAA - STAT. All four ERTS-1 MSS channels were used to classify the study area using the LARSAA - CLASSIFY processor. The result was that only water which was spectrally similar to Lake Somerville was displayed as water and all of the small ponds, even those which were used as training fields, were not statistically represented in the thresholding process. Thresholding was necessary because, when only water training fields are used, every pixel in the scene is classified as water and the threshold must be used to determine those pixels which are close to the water class mean vector in spectral space.

In a second attempt to employ the LARSAA - CLASSIFY processor, the training fields were selected to give equal weight to large, clear lakes and to small, turbid ponds. A training field of 8 pixels was selected from Lake Somerville to represent clear water and a training field of 8 pixels was selected from a turbid pond downstream from Lake Somerville. The turbid pond was selected and identified using the

channel 4 grey map and the aircraft photography. The channel 4 data values were known before the lakes were chosen, but were not known for channels 1, 2, or 3. The two training fields were combined to form a class called "water". Three other classes were included in order to induce the processor to fix interclass boundaries in the spectral space, rather than classifying all pixels as water (which is what would happen if only the water training fields were used). The other three classes were areas of high soil moisture, trees and a high spectral reflectance area (probably bare soil). The exact location in spectral space of the other classes is not important as long as they occur in a discrete part of the space relative to the location of water. The classes were chosen by looking for large, homogeneous non-water areas in the high altitude aircraft photography and they trying to cover those areas with blocks of pixels in the ERTS-1 data.

The mean vector for the water class was computed to be
Chan 1 - 28.94
Chan 2 - 24.38
Chan 3 - 16.50
Chan 4 - 3.50

The covariance matrix for the water class was

80.99			
122.96	188.25		
94.63	144.27	113.60	
19.63	29.86	32.60	5.20

The diagonal elements are the variances. They are large because the two training fields are widely separated in spectral space. The off-diagonal elements are the covariances, and they are large because the data are highly correlated. The covariance must be no greater in absolute value than the geometric mean of the two associated variances. Thus, for example, σ_{14}^2 which is equal to 19.63, could never exceed

$\sqrt{80.99 \times 5.20}$ = 20.52 and it is seen to be quite near its maximum value. Thus, all points near the line in spectral space joining the data representing the two training fields will be classified as water. Also, points near the extension of this line beyond the data which represents the two training fields will be classified as water. The use of this water class, plus the three arbitrary non-water classes resulted in a correct identification of all water bodies of 10 surface acres or greater in the study area (a total of 12 lakes). It should be noted that the frequency of false detections was zero.

A parallel effort was pursued to extract statistics for water training data from the selected ERTS-1 scene using the ISOCLS program instead of manual selection of the training fields, and the generation of the statistics using LARSAA - STAT processor. Data processing runs were made using the modified ISOCLS program to cluster data related to the selected study area and the allowable cluster dimensions in spectral space were reduced until the data related to small bodies of water were separated into different clusters. Eleven of the 29 clusters obtained were correlated to surface water. The ISOCLS program produces a statistics deck in the format compatible with the LARSAA - CLASSIFY processor. Thus, a representative area may be clustered and the results may be used to classify a larger area. The procedure still requires manual intervention to identify which clusters represent water and to decide if the spectral space has been partitioned finely enough to separate all of the water of interest. Also, when the number of classes becomes large, the running time for the programs becomes significantly greater. The ISOCLS program required more than one hour to stabilize on the Univac 1108 for 1/16 of an ERTS-1 strip (1/64 of a scene). The LARSAA processor takes 9 hours to classify and display one ERTS-1 strip (1/4 of a scene) when using 29 classes. the LARSAA program does a large amount of tape handling when classifying a large area, and it was found that it was improbable to have a run go to completion because of system errors which occurred while trying to process data tapes as often as this volume of data required.

Because previous experience had indicated a good potential for surface water mapping using only channel 4 grey maps (see discussion on PICMON, Section 3.2.1), an effort was made to use that single channel in an automated procedure which would not require manual intervention. However, a check of some of the working grey maps revealed that pixels composed of mixtures of water and non-water gave the same data value in channel 4 as some of the pixels representing areas of surface water. Therefore, use of a channel 4 discriminant above resulted in less than satisfactory performance, and some additional information would be needed in such an automated technique. The 29 cluster means from the ISOCLS run were plotted to see if the inclusion of one more channel of data would enhance the ability to discriminate between standing water and persistent surface water. The data from the other three channels was plotted against channel 4 data and the resultant eleven water clusters were identified. The Channel 1/Channel 4 plot presented in Figure 2 showed the most promise of successful discrimination because the water clusters were better separated from the other clusters with these two channels relative to the other two plots that were generated. The simplest method of separating water from non-water on the two channel plot was to draw a straight line with all of the water clusters on one side of the line and all of the non-water clusters on the other. Two lines each passing through the origin, were used that had slightly different slopes to segregate the data into water and non-water regions. A comparison was made with the airborne photographic "ground truth" data.

It is possible to create class means and covariance matrices which place a straight line interclass boundary in the Channel 1/Channel 4 spectral space using LARSAA. By using the same distance measure in two classes and by giving the classes circular symmetry, the interclass boundary, which is equidistant from both class means, must be a straight line. The location and orientation of the straight line is determined by the locations of the two class means. Two straight lines through the

origin were tested for performance (Figure 2). The slopes of the lines were chosen to avoid passing through any integer values which could result in unpredictable and unacceptable performance results. The first line, with a slope of 3.727, classified pixels as water when in reality these pixels were mixture pixels containing portions of surface water and other cover material. The second line, with a slope of 4.556, eliminated false detections related to the mixture pixels, but failed to correctly identify pixels related to areas of turbid water. The two hypothetical class means were then chosen to be symmetrically located on opposite sides of the desired linear decision boundary. Thus, the intersection of the linear discriminant line and the line joining the means of the two classes was placed near the bulk of the plotted data. It avoided the possibility of having the LARSAA - CLASSIFY processor generate large negative numbers as it computes the logarithm of the classification probabilities. The covariance matrix was taken to be

2

0   2

0   0   2

0   0   0   2

causing the unit of distance to be $\sqrt{2}$ thus keeping the numbers within a tractable range for the computer. The fact that the off-diagonal elements are zero indicates that there is no correlation between channels, and that the classes have circular symmetry. The curves of constant distance from the class means are circles in the two dimensional spectral space of Channel 1/Channel 4. The same covariance matrix was used for both classes in all of the straight line studies.

The third procedure employed a line which did not pass through the origin (see Figure 2). This line passed through the two points (0, 8.5) and (11.5, 41) with a slope of 2.826. This line resulted in correct detection and identification of all water surfaces of 10 surface acres or greater in the study area, and resulted in a zero frequency of false detections of surface water of 10 surface acres or greater.
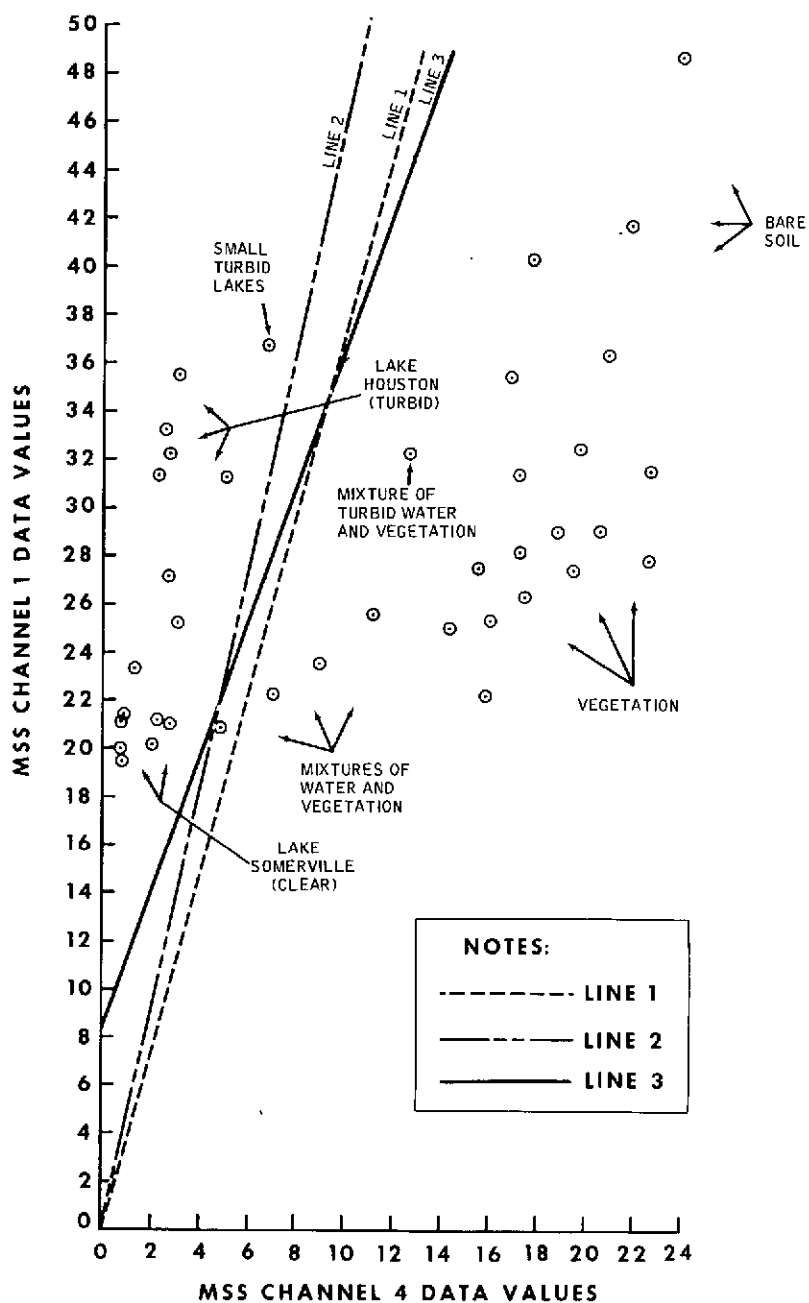
Figure 2. — Location of cluster means in a Channel 1/ Channel 4 spectral space. The three attempts to divide water means from contending means are indicated by lines 1, 2, and 3.

4-8

The procedure used to obtain a linear, two-channel decision boundary using the LARSAA processor has been described in this section. This procedure will be described, step-by-step, below in order to document the method of using the LARSAA processor to generate a linear discriminant boundary.

The line to be tested passes through the (0,8.5) and (11.5, 41) points in the Channel 1/Channel 4 spectral space where the first element of each vector represents the Channel 4 data value. The midpoint of the line joining those two points is (5.75, 24.75). The slope of the line is (41-8.5)/11.5 = 65/23. The two class means must lie on the perpendicular bisector of the line. The perpendicular bisector passes through the midpoint of the line and has a slope of -23/65.

The two class means must be on the perpendicular bisector at equal distances on opposite sides of the midpoint (5.75, 24.75). Their distance from the midpoint is arbitrary and was selected to be 3 units away in the Channel 4 direction. Thus, the Channel 4 values for the two class means were (5.75 - 3) and (5.75 + 3), or 2.75 for the water class and 8.75 for the non-water class.

The channel 1 values for the class means are then computed by substituting the channel 4 values into the equation for the perpendicular bisector. A shortcut was actually used as follows. The channel 1 displacement from the midpoint is the channel 4 displacement multiplied by the slope of the perpendicular bisector or (-3) x (-23/65) for water and (3) x (-23/65) for non-water which is 1.061538 except for the sign. The channel 1 values for the class means are, therefore, 24.75 + 1.061538 = 25.811538 for water and 24.75 - 1.061538 = 23.688462 for non-water.

It is necessary to put some values into the processor for channels 2 and 3 in order to force the card reading subroutine to count to 4 and place the channel 4 data values in the correct memory locations. Values

of 14.0 for channel 2 and 8.0 for channel 3 were selected as filler numbers but they were made the same in both classes as a precaution. If it became necessary to process all four channels or if the FEATURES card was inadvertently forgotten and all four channels were used by default, the equality of the Channel 2 and Channel 3 data values would cause the class boundary hyperplane to be perpendicular to the Channel 1/Channel 4 plane and to cut it in the desired line. This insures the degenerate case and the result is independent of whether or not channels 2 and 3 are used. When the computer run was made, only channels 1 and 4 were actually used and the running time for LARSAA - CLASSIFY was about 30 percent of what would have been necessary for the case of all 4 channels.

The class means thus generated are:

water (25.811538, 14.0, 8.0, 2.75)

non-water (23.688462, 14.0, 8.0, 8.75)

The covariance matrices are:

2

0   2

0   0   2

0   0   0   2

for both classes. It is necessary to load the full covariance matrix in order to get the channel 1 and channel 4 into the proper memory locations, even though only 3 of the 10 elements will actually be used. The channel 2 and 3 variances are also set equal to 2 although their actual values are almost immaterial.

Figure 3 shows a black-and-white rendition of a color infrared image (from aircraft Mission 220) of a portion of the study area containing inundated fields and small ponds. The areas whose data are near the straight line decision boundary are indicated and are labelled as fields (F) or lakes (L). The tabulation below gives the surface areas of the lakes which were detected or missed depending upon the
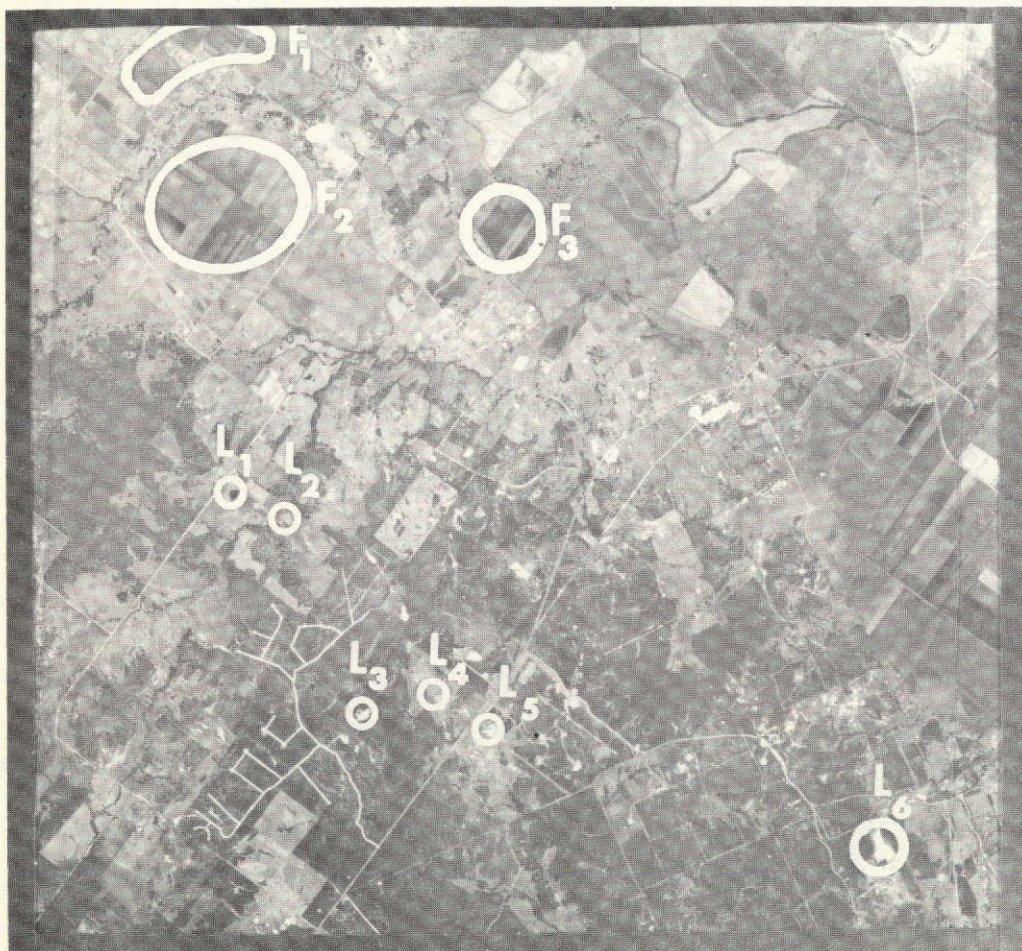
4-10

Figure 3. — Black and white reproduction of color infrared
image denoting a portion of the study area. The F's refer
to cultivated fields in the image. The L's refer to lakes
in the image which were detected or lost according to the
positions of the lines shown in Figure 2. The F and L
designations are maintained through Figures 4, 5, and 6.

location of the straight line discriminant in two-dimensional spectral space.

L_1 3.0 acres

L_2 5.4 acres

L_3 3.2 acres

L_4 4.8 acres

L_5 7.2 acres

Figures 4, 5, and 6 are line printer classification maps of the area shown in the photograph in Figure 3. Figure 4 is the map resulting from the use of a linear discriminant with a slope of 3.727 (line 1) and shows that mixture pixels were classified as water. Figure 5 is a map resulting from the use of the line with a slope of 4.556 and shows that areas related to turbid water were incorrectly classified as non-water. Figure 6 is a map as a result of using the line with an intercept of the vertical axis of 8.5 and a slope of 2.826. It shows that mixture pixels were classified as non-water, in contrast to the results obtained and shown in Figure 4. In addition, the pixels related to turbid water were correctly identified, in contrast to the results shown in Figure 5. This linear decision boundary with an intercept of 8.5 and slope of 2.826 provided the basis for a recommended discriminant function for an operational data processing system.

Figure 7 shows a channel 1/channel 4 histogram of the data for the ERTS-1 scene containing the study area. Such histograms have been found to be useful for visualizing the separation of the water data from the data for the rest of the scene. The stepped line in the histogram represents the linear decision boundary as implemented on the production processor. If any of the integer pairs in the histogram are identified as contributing a large number of false alarms, they may be deleted by changing the tabular value in the processor. The appropriate values can then be input with data cards.
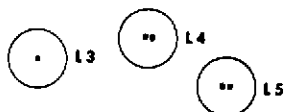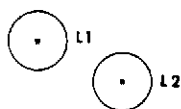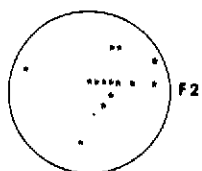
Figure 4. — Computer Line Printer output for the performance
evaluation of the first line with a slope of 3.727. The
larger lake, L6, is detected together with the false detec-
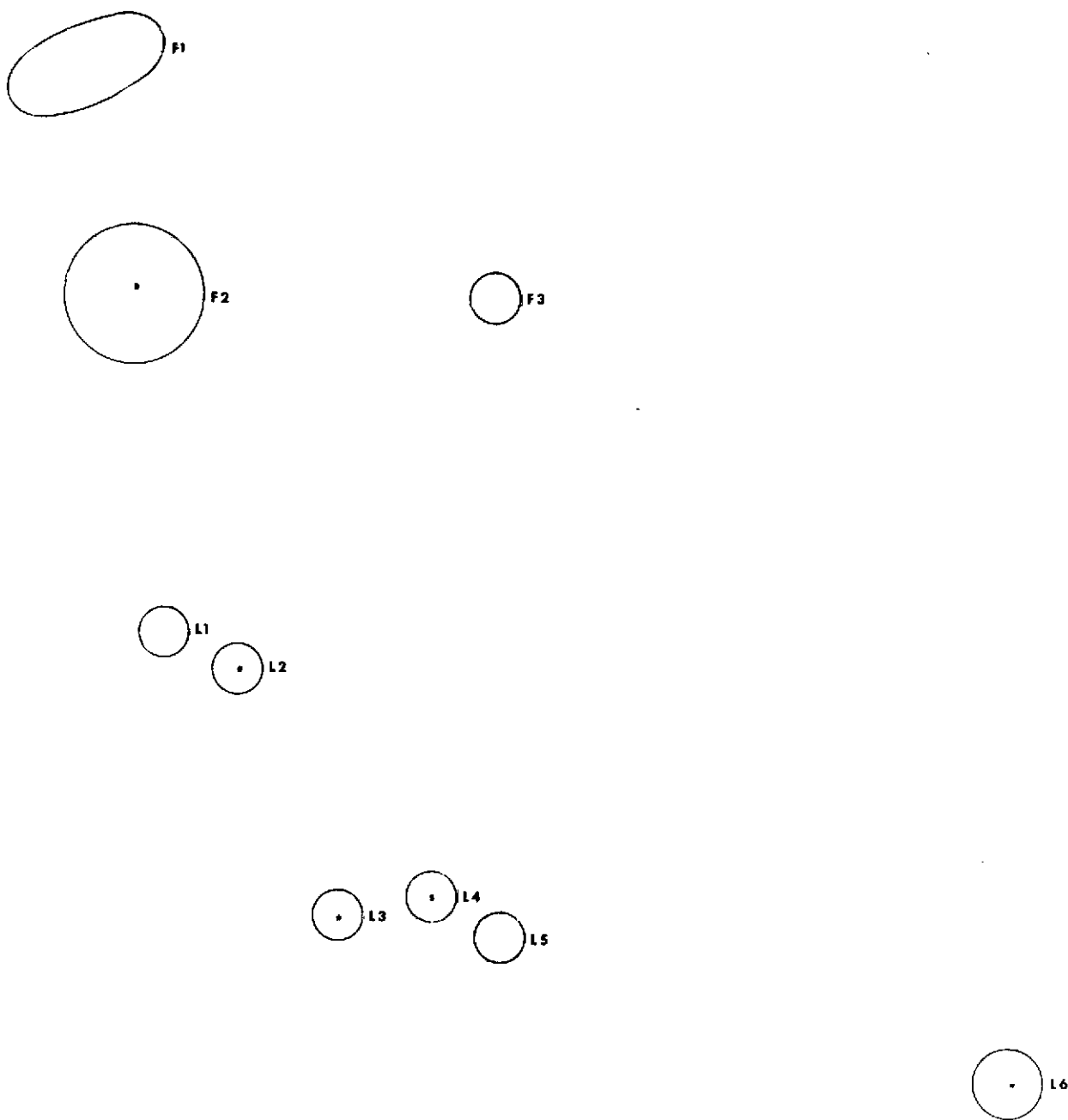tions in the three inundated fields.

4-13

Figure 5. — Computer Line Printer output for the performance evaluation of the second line with a slope of 4.556. The false detections in the inundated fields are reduced to one pixel, but the larger lake, L6, is reduced to one pixel and, therefore, is not identified as a larger lake.
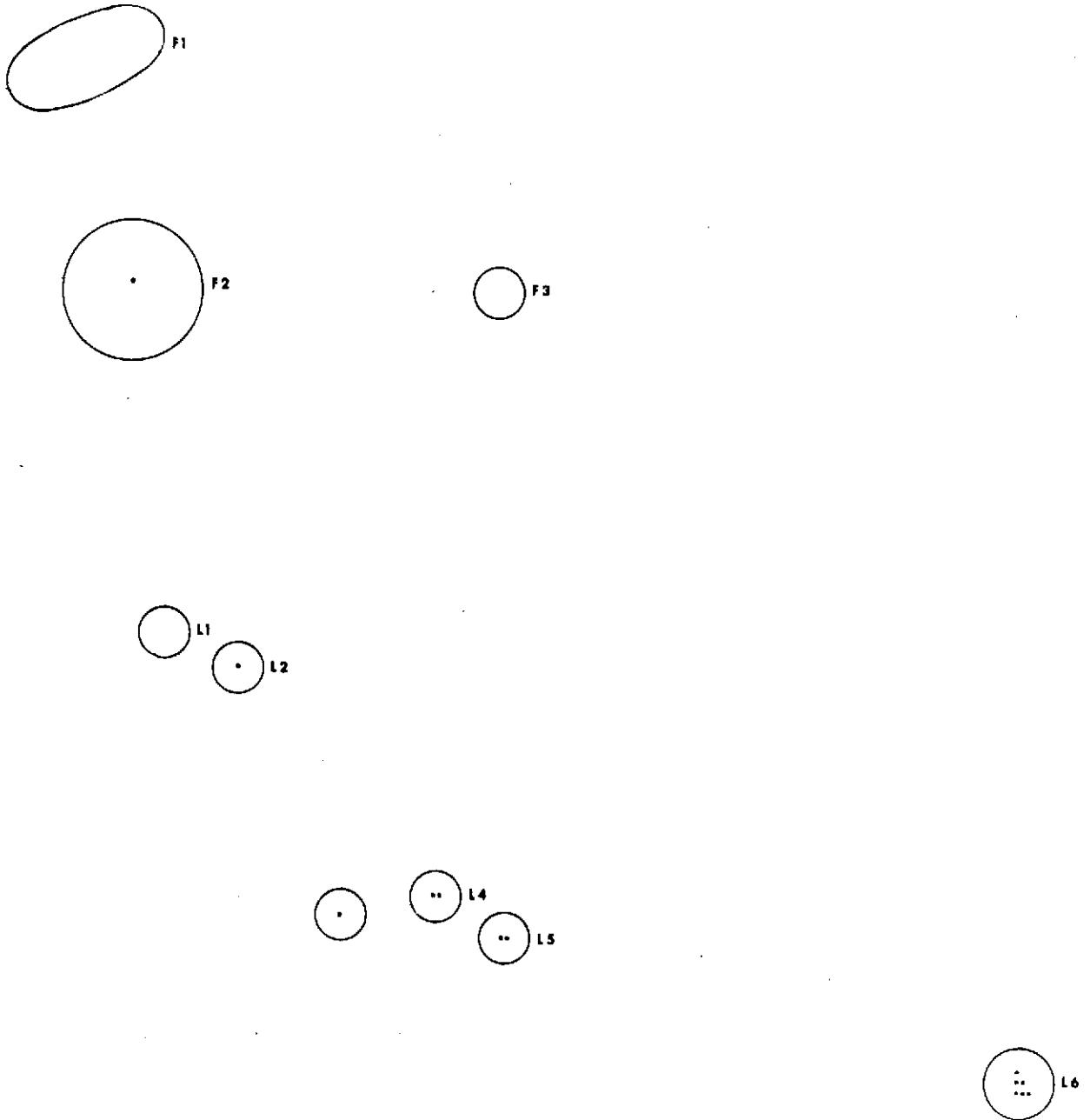
Figure 6. — Computer Line Printer output for the performance
evaluation of the third straight line with a slope of 2.826
and an intercept the channel 1 axis of 8.5. The false
detections originally found in the inundated fields continue
to be eliminated and the larger lake, L6, is detected.

4-16

CHANNEL 4 DATA VALUES ACROSS TOP
CHANNEL 1 DATA VALUES DOWN EDGE
FRAME 1092-16305 - STRIP 3 - 23 OCT 72
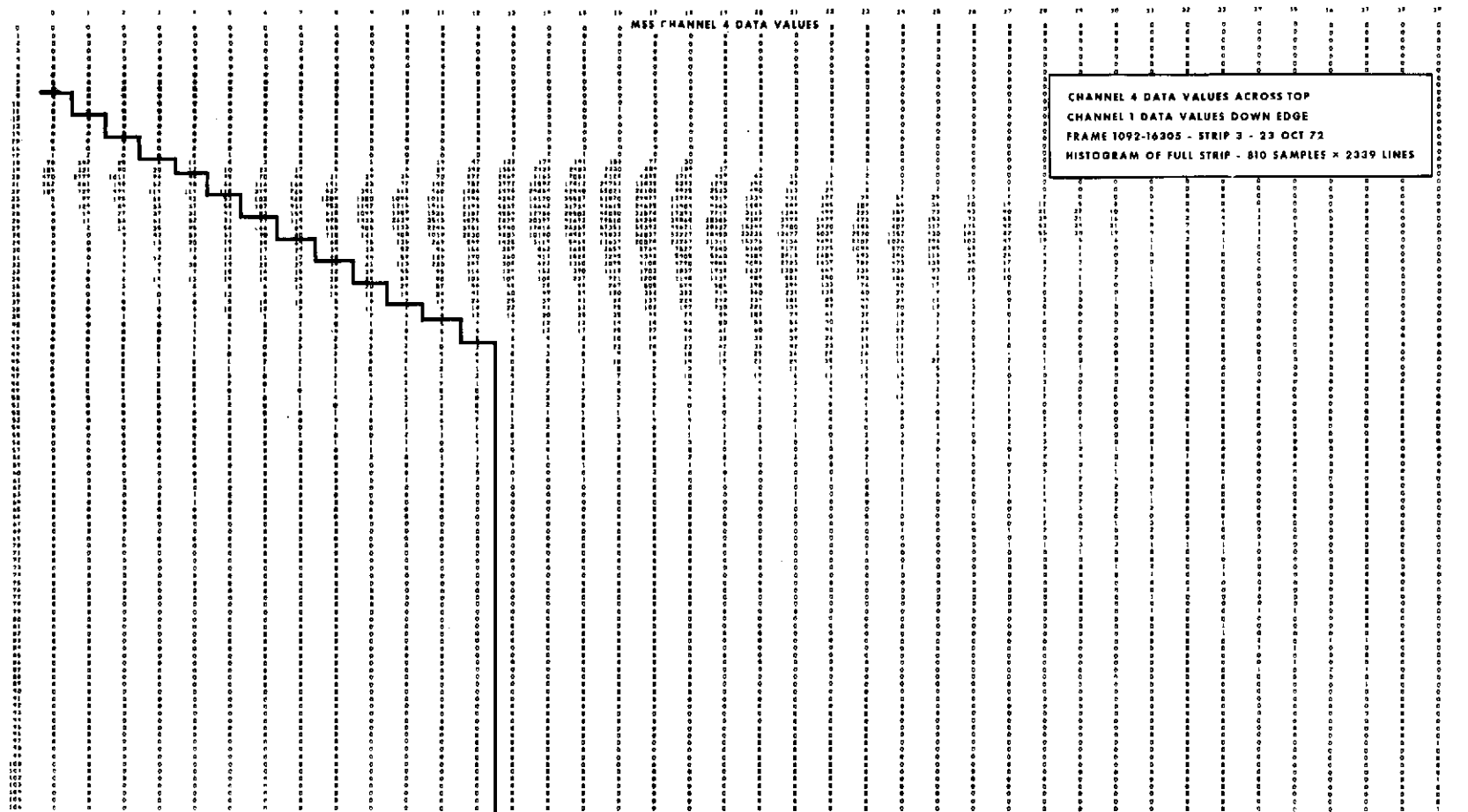HISTOGRAM OF FULL STRIP - 810 SAMPLES × 2339 LINES

Figure 7. - Two channel histogram of the complete ERTS strip which contains the study area. The decision boundary which was implemented in the production processor is shown as a stepped line. It gives the same result as the straight line in the LARSAA processor.

5.0 CONCLUSIONS AND OPERATING CONSTRAINTS

The linear interclass boundary between water and non-water which has been described in this report resulted in 100 percent correct detection and identification of all the surface water areas of 10 surface acres or greater in the study area in addition to a zero frequency of false detections of all water areas of 10 surface acres or greater. The use of a linear decision boundary for different ERTS-1 scenes requires that the water data be well separated from the data for the rest of the scene. It also requires that the water data be stable and relatively invariant from scene to scene. Both of these conditions are satisfied by the water data for study areas in east Texas. The variations which occur in water data are due to two factors: (1) the inclusion of vegetation in the scanner field-of-view along with the water causes the channel 4 data value to increase with relatively little increase in the channel 1 data value and, (2) an increase in the turbidity of the water causes the channel 1 data value to increase with relatively little increase in the channel 4 data value. These two sources of variability are well handled by the recommended linear decision boundary in two dimensional spectral space.

The rate of occurrence of water in the data studied was considerably less than 1 percent. Additional data were not available for improving the location of the line in spectral space or for studying the possibility of using curves more general than a straight line. In fact, false detections using the linear discriminant function have occurred in subsequent data processing efforts. These false detections have been related to pixels located in areas of cloud shadows and terrain shadows.

The sun elevation factor was determined to have little effect on the data values for water, but it produces significant effect on the data values for non-water surface cover in the scene. As the sun elevation decreases from July to December (60° to 30°), the data values

related to non-water areas in the scene approach the water pixel data
values. Decreased performance using the recommended linear discriminant
may occur if this shift in non-water pixels in spectral space is suffi-
cient to translate the non-water pixels into the water region of spectral
space. The frequency of false detections related to terrain shadows will
potentially increase when using data taken at the lower sun elevation
angles. These difficulties may be minimized by a judicious selection
of ERTS-1 scenes.

Atmospheric effects may potentially affect the correct identifica-
tion of water pixels, as well as the frequency of false detections.